JASA JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

december 1972 volume 67 number 340

ADVERTISING: Write to Assistant to the Editor, University of Illinois, Chicago Circle, Box 6905, Chicago, Illinois 60680.

SUBSCRIPTIONS: Send to American Statistical Association, 806 15th Street, N. W., Washington, D. C. 20005.

RATES: $25.00 per year worldwide. Current single copy price, $6.50. Prices for back issues available on request.

CHANGE OF ADDRESS: Send new address with zip code and address label from JASA to American Statistical Association, 806 15th Street, N. W., Washington, D. C. 20005. Allow eight weeks.

PUBLISHED: Quarterly by the American Statistical Association. Business Office: 806 15th Street, N. W., Washington, D. C. 20005. Printed by George Banta Co., Inc., Menasha, Wisconsin, Editorial Office: University of Illinois, Chicago Circle, Box 6905, Chicago, Illinois 60680. Second class postage paid at Washington, D. C. and at additional mailing offices.

Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis

D. B. DUNCAN and S. D. HORN*

A large class of useful multivariate recursive time series models and estimation methods has appeared in the engineering literature. Despite the interest and utility which this recursive work has when viewed as an extension of regression analysis, little of it has reached statisticians working in regression. To overcome this we (a) present the relevant random-β regression theory as a natural extension of conventional fixed-β regression theory and (b) derive the optimal recursive estimators in terms of the extended regression theory for a typical form of the recursive model. This also opens the way for further developments in recursive estimation, which are more tractable in the regression approach and will be presented in future papers.

1. INTRODUCTION

A very useful class of multivariate time series models and estimation methods has been developed in the engineering literature (by Kalman [5, 6] and others) under titles such as *linear dynamic recursive filtering and prediction*. These are useful for all sorts of problems varying from missile trajectory estimation to estimating and predicting the condition of a critically ill medical patient.

Despite a basic closeness which this work has to classical regression theory and despite its high utility for providing simple solutions to otherwise difficult extensions of regression problems, little of this work has come to the attention of the average statistician working in regression theory. We feel this lack of communication is due to the fact that the new theory has not been developed as a direct extension of regression theory, but instead from a different, though elegant, point of view of wide sense conditional probability theory (e.g., [3]).

In an effort to overcome this communications block we show how a typical form of linear dynamic recursive estimation can be developed as a natural extension of regression theory. To do this, we must present first a wide sense random-β regression theory of considerable interest in its own right. Some of these ideas have appeared previously in the statistical literature (e.g., [2], [4], [7], [9]).

The Kalman time series models are expressed in terms of two recursive equations. A typical form for a p-variate time series $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$ consists of (1) an observation equation: $\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$ for times $t = 1, \cdots, n$ and (2) a dynamic regression coefficients transition equation: $\boldsymbol{\beta}_t = \boldsymbol{T}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{u}_t$, $t = 2, \cdots, n$, starting with $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_1 + \boldsymbol{u}_1$ at $t = 1$ where $\boldsymbol{\mu}_1$ is a known prior mean for $\boldsymbol{\beta}_1$. Each observation equation is of the familiar regression form in which the regression coefficient vector $\boldsymbol{\beta}_t(r \times 1)$ is called the *state* at time t, $\boldsymbol{X}_t(p \times r)$ is a known matrix of regressors, and $\boldsymbol{\varepsilon}_t(p \times 1)$ is a vector of random errors.

The dynamic state transition equation $\boldsymbol{\beta}_t = \boldsymbol{T}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{u}_t$ is new to conventional fixed-β regression theory and expresses the state $\boldsymbol{\beta}_t$ as a known linear transformation $\boldsymbol{T}_t(r \times r)$ of the previous state $\boldsymbol{\beta}_{t-1}$ plus a vector $\boldsymbol{u}_t(r \times 1)$ of random errors.

The linear operation on $\boldsymbol{\beta}_{t-1}$ and the introduction of randomness are what give $\boldsymbol{\beta}_t$ its *linear dynamic* aspect. A situation where this is of particular value, e.g., is in the modelling of a "trajectory" (interpreted broadly as any smooth time function) in which the parameters describing the trajectory at time t depend on those at time $t-1$.

The dynamic state transition equations, the starting equation $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_1 + \boldsymbol{u}_1$, and assumptions about the error vectors $\boldsymbol{u}_t$, $t = 1, \cdots, n$ provide a prior model for the complete vector $\boldsymbol{\beta}$ of all the regression coefficients or state elements concerned.

As in conventional regression theory, the assumptions made about the error vectors $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are of one of two forms. In the stronger form, the *Gaussian* assumptions, all the error vectors have zero expectations, all are uncorrelated with one another, and all have known variance-covariance matrices $\boldsymbol{Q}_t = \mathrm{var}(\boldsymbol{u}_t)$ and $\boldsymbol{R}_t = \mathrm{var}(\boldsymbol{\varepsilon}_t)$ for all t. Each error vector has the multivariate normal (Gaussian) distribution, $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_t)$. In the weaker form, the *wide-sense* assumptions, no distributional assumptions are made except the same ones specifying the first and second moments. We denote these by $\boldsymbol{u}_t \sim WS(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $\boldsymbol{\varepsilon}_t \sim WS(\boldsymbol{0}, \boldsymbol{R}_t)$. In the Gaussian case, of course, all the error vectors are independent.

The following specializations reduce the new dynamic model to the conventional fixed-β regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$: (1) The state transition equation $\boldsymbol{\beta}_t = \boldsymbol{T}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{u}_t$ is now the trivial identity transformation (i.e., $\boldsymbol{T}_t \equiv \boldsymbol{I}$ and $\boldsymbol{u}_t \equiv \boldsymbol{0} \leftrightarrow \boldsymbol{Q}_t \equiv \boldsymbol{0}$) and $\boldsymbol{\beta}_1$ is fixed and unknown. This is the only vector of regression coefficients involved, so the need for the subscript is removed. Thus we may write $\boldsymbol{\beta}_t = \boldsymbol{\beta}_1 = \boldsymbol{\beta}$ for all t. (2) The tth observation equation

* D.B. Duncan is professor and S.D. Horn is assistant professor, both with the Departments of Biostatistics and Mathematical Sciences, Johns Hopkins University. Research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under contract number F44620-70-C-0066. Technical Report No. 162 of the Department of Statistics; Publication No. 478 of the Department of Biostatistics.

© Journal of the American Statistical Association
December 1972, Volume 67, Number 340
Theory and Methods Section

$y_t = X_t\beta_t + \varepsilon_t$ is univariate $p=1$ and represents the tth row $y_t = x_t'\beta + \epsilon_t$ of the conventional model $y = X\beta + \varepsilon$ where $y_t(1\times 1)$ is the tth element of y, $x_t' = X_t(1\times r)$ is the tth row of X, etc.

The most important advantage of the new work is the *increased versatility* of the dynamic model, i.e., the enlarged range of real and important problems involving univariate or multivariate $(p \geq 1)$ non-stationary time series with which it can cope.

Another advantage is the *simplicity* of the recursive methods for the efficient handling of the optimal estimation of extremely large numbers of parameters. A sequence of very simple recursive solutions replaces the direct simultaneous solution for all of the β parameters involved. In a typical form of these methods, filtering, an optimal estimator b_t of β_t is obtained recursively as a linear combination of the previous estimator b_{t-1} and of y_t, the response vector at time t. The recursive filtering equations simultaneously provide the variance-covariance matrix $S_t = \text{var}(b_t - \beta_t)$ of the estimator. Both equations have been developed in an updating form which is useful especially in programmed computing (cf. (3.5a) and (3.5b)).

It is well known from the engineering literature that under the wide sense error assumptions the filter estimator b_t for β_t is optimal in the sense of being the minimum mean square linear estimator based on all the data $\mu_1, y_1, \cdots, y_t$ up through time t. Under the Gaussian assumptions it is the minimum mean square estimator. Variations of the updating equations give similarly optimal estimators $b_{t|s}$ of β_t based on all the data $\mu_1, y_1, \cdots, y_s$ up through time s, for varying values of s and t. Of special interest are the estimators $b_{t|n}$ which are optimal based on all the data.

Despite its advantages as a natural extension of conventional regression theory, the engineering approach has so far been little understood or used by the conventional regression analyst, perhaps due partly to the unfamiliarity of the notation and terms in the engineering literature. But more importantly, the neglect is due to a confusion caused by the use, in its development, of a theory in which the names are the same as in conventional regression theory but the roles are different. Specifically, the central objective in both cases is to estimate β, the entire vector of regression coefficients involved. Speaking in terms of vector space projections, as Kalman does, the conventional regressionist approaches this problem by projecting an observation y vector onto a linear vector space spanned by the columns of an X matrix, i.e., by "fitting a regression $\hat{y}$ of y on X." The optimal estimates b of β are then found as the coefficients of the orthogonal projection $\hat{y}$ so obtained, i.e., the coefficients of X in the equation for $\hat{y}$. The Kalman [6] approach, on the other hand, makes use of the fact that β_t is now a random variable. By a suitable redefinition of the vector space terms involved, the optimal estimate *itself* is found directly as an orthogonal projection $b_{t|s}$ of β_t on a vector space spanned by the vectors $\mu_1, y_1, \cdots, y_s$ of the data involved. In regression terms, this is "fitting a regression of β_t on $\mu_1, y_1, \cdots, y_s$." The normal equations for doing this are based on the *expected* first and second moments of the variables involved.

The purpose of the present paper is to overcome the communications block by: (1) explaining the new dynamic models and methods in the terms and notation of conventional regression theory, and (2) deriving the optimality of the recursive estimators in terms of a natural random-β extension of the conventional fixed-β regression theory.

The random-β regression theory approach should serve as a two-way communications bridge; a bridge by which the conventional regressionist can move to an easier understanding of the more direct derivation of recursive estimators, and by which the engineer can move to an easier understanding of the relevance to his problems of many valuable results in regression theory. Apart from these objectives, the random-β extensions of fixed-β regression theory, e.g., the extended Gauss-Markov theorem, are of considerable interest in themselves, and of potential value for a wide class of applications.

2. A WIDE-SENSE RANDOM-β REGRESSION THEORY

In this section we extend well-known definitions and theorems of fixed-β regression theory in a natural way to form a random-β regression theory required for a simple regression development of the Kalman recursive equations.

2.1 The Model

The model can be written in the familiar fixed-β regression form

$$y = X\beta + \varepsilon \quad \text{where} \quad \varepsilon \sim WS(0, \Sigma) \tag{2.1}$$

except that β now is random and may be written in a similar style as

$$\beta = \mu + u \quad \text{where} \quad u \sim WS(0, \Sigma_\beta) \tag{2.2}$$

and u is uncorrelated with ε. The notation $z \sim WS(\mu, \Sigma)$, as used in the introduction, indicates the wide sense model that z has any distribution with mean μ and variance Σ.

The dimensions and properties of the vectors and matrices involved are:

- y is an $n\times 1$ vector of random variables or observations on same,
- X is an $n\times r$ matrix of known regressors,
- β is an $r\times 1$ vector of unknown parameters (regression coefficients),
- ε is an $n\times 1$ vector of observation errors,
- μ is an $r\times 1$ known prior mean for β,
- u is an $r\times 1$ vector of prior errors,
- $\Sigma(n\times n)$ and $\Sigma_\beta(r\times r)$ are variance matrices both of which are known except possibly for a scale factor σ^2.

The Condensed Model. By writing

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon} \tag{2.3}$$

for

$$\begin{bmatrix} \mu \\ y \end{bmatrix} = \begin{bmatrix} I \\ X \end{bmatrix} \beta + \begin{bmatrix} -u \\ \varepsilon \end{bmatrix}$$

the whole model, (2.1) and (2.2) above, may be written more briefly as

$$(\tilde{y} - \tilde{X}\beta) \sim WS(0, \tilde{\Sigma}) \tag{2.4}$$

where

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma \end{bmatrix} \tag{2.5}$$

with dimensions $\tilde{y}(\tilde{n}\times 1)$, $\tilde{X}(\tilde{n}\times r)$, $\tilde{\Sigma}(\tilde{n}\times\tilde{n})$, $\tilde{n}\equiv r+n$.

It is precisely this condensation that allows us to extend the familiar fixed-β regression theory to the random-β case with a minimum of new techniques needed to prove the required results.

In fixed-β theory an important estimator is the WLSE (weighted least squares estimator) of β. The natural extension of this in random-β theory is:

Definition 2.2: Weighted Least Squares Estimator. The vector $b(r\times 1)$ is the WLSE of β if

$$(\tilde{y} - \tilde{X}\hat{\beta})'\tilde{\Sigma}^{-1}(\tilde{y} - \tilde{X}\hat{\beta}) \tag{2.6}$$

is minimized when $\hat{\beta} = b$.

Theorem 2.3: Normal Equations. The WLSE of β is given by the solution of the Normal Equations

$$(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})b = \tilde{X}'\tilde{\Sigma}^{-1}\tilde{y}. \tag{2.7}$$

Proof: The quadratic form (2.6) to be minimized has the same form as that in fixed-β theory where it is known that the minimum is given by the solution of (2.7).

By substitution the estimator b given by (2.7) is seen to be the familiar weighted combination:

$$b = V[\Sigma_\beta^{-1}\mu + X'\Sigma^{-1}y] \tag{2.8}$$

where

$$V = [\Sigma_\beta^{-1} + X'\Sigma^{-1}X]^{-1} \tag{2.9}$$

The distribution of the WLSE b of β in wide-sense fixed-β theory is known to be $b \sim WS(\beta, (X'\Sigma^{-1}X)^{-1})$. Note that it is this fixed-β estimator weighted inversely by its variance which gives the $X'\Sigma^{-1}y$ term in (2.8). Intuitively, in the random-β case interest is centered on the distribution of $b-\beta$, since this difference tells how well a random β is being estimated.

Lemma 2.4: Wide-Sense Distribution of $b-\beta$. Given that

$$(\tilde{y} - \tilde{X}\beta) \sim WS(0, \tilde{\Sigma})$$

and

$$b = (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}\tilde{y}$$

then

(1) $(b - \beta) \sim WS(0, (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1})$

(2) $E(b - \beta)(\tilde{y} - \tilde{X}b)' = 0$, and

(3) $E(b - \beta)\tilde{y}' = 0$.

Proof: We first note that

$$\begin{aligned}(b - \beta) &= (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}(\tilde{y} - \tilde{X}\beta) \\ &= (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}\tilde{\Sigma}^{-1}\tilde{\varepsilon};\end{aligned} \tag{2.10}$$

$$\begin{aligned}(\tilde{y} - \tilde{X}b) &= [I - \tilde{X}(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}]\tilde{y} \\ &= [I - \tilde{X}(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}](\tilde{X}\beta + \tilde{\varepsilon}) \\ &= [I - \tilde{X}(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}]\tilde{\varepsilon}.\end{aligned} \tag{2.11}$$

Results (1) and (2) follow directly.

From (2.1) and (2.2)

$$y = X\beta + \varepsilon = X\mu + Xu + \varepsilon. \tag{2.12}$$

From (2.10) and (2.12) we get

$$\begin{aligned}E(b-\beta)y' &= (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1} \\ &\cdot E\left\{\begin{bmatrix} -u \\ \varepsilon \end{bmatrix}\mu'X' + \begin{bmatrix} -u \\ \varepsilon \end{bmatrix}(u'X' + \varepsilon')\right\} \\ &= 0 + (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}(-\Sigma_\beta^{-1}\Sigma_\beta X' + X'\Sigma^{-1}\Sigma) = 0.\end{aligned}$$

Since $E(b-\beta)\mu' = 0$ also, result (3) follows.

Definitions 2.5: An estimator $\hat{\gamma}(q\times 1)$ of any $q\times 1$ linear vector function $\gamma = C\beta$ of β is

1. *linear* if it is linear in $\tilde{y}$, i.e., if $\hat{\gamma} = A\tilde{y} = A_1\mu + A_2y$, where A, A_1 and A_2 are $q\times\tilde{n}$, $q\times r$ and $q\times n$ respectively;
2. *unconditionally unbiased* (*u-unbiased*) if $E(\hat{\gamma} - \gamma) = 0$;
3. *minimum mean square* if $E(\hat{\gamma}_i - \gamma_i)^2$ is minimized for each $i = 1, \cdots, q$;
4. *minimum variance* if var $(\hat{\gamma}_i - \gamma_i)$ is minimized for each $i = 1, \cdots, q$;
5. *the MVLUE* (minimum variance linear u-unbiased estimator) if for every $\hat{\gamma}_i$ no other estimator has equal or smaller variance in the class of all linear u-unbiased estimators of γ_i, $i = 1, \cdots, q$;
6. *the MMSLE* (minimum mean square linear estimator) if for every $\hat{\gamma}_i$ no other estimator has equal or smaller mean square in the class of all linear estimators of γ_i, $i = 1, \cdots, q$.

Because a linear estimator $\hat{\gamma}$ is linear in μ as well as y we get the following useful result in the random-β case.

Lemma 2.6: Equivalence of the MVLUE and the MMSLE. Given that $(\tilde{y} - \tilde{X}\beta) \sim WS(0, \tilde{\Sigma})$, the MVLUE of any linear vector function $\gamma = C\beta$ of β is the MMSLE of γ.

Proof: Suppose $\mu \neq 0$. Let $\hat{\gamma}$ be linear, i.e., $\hat{\gamma} = A_1\mu + A_2y$, and consider

$$E(\hat{\gamma}_i - \gamma_i)^2 = [E(\hat{\gamma}_i - \gamma_i)]^2 + \mathrm{var}(\hat{\gamma}_i - \gamma_i) \quad \text{for } i = 1, \cdots, q. \tag{2.13}$$

Since A_1 and μ are non-random, the var$(\hat{\gamma}_i - \gamma_i)$ does not

depend on A_1 (or $\mathbf{u}$) for $i=1, \cdots, q$ and we can minimize the right-hand side of (2.13) in two steps:

1. Choose A_2 to minimize var $(\hat{\gamma}_i-\gamma_i)$ for $i=1, \cdots, q$, and then
2. choose A_1 (which will be a function of A_2) so as to make the bias $E(\hat{\gamma}_i-\gamma_i)=0$ for $i=1, \cdots, q$.

From Step 2 it follows that the MMSLE is u-unbiased and thus the MVLUE is the MMSLE of γ. If $\mathbf{u}=\mathbf{0}$ then $E(\hat{\gamma}-\gamma)=\mathbf{0}$ and we see immediately from (2.13) that the MVLUE is the MMSLE of γ.

We are now ready to prove the natural random-β extension of the Gauss-Markov theorem. We prove the theorem by a well-known style like that of Bose [1] which ascribes a property that may be called "wide-sense sufficiency" to the right-hand side of the normal equations for estimating any linear vector function $\gamma=C\beta$ of β.

Theorem 2.7: The Extended Gauss-Markov Theorem. Given that $(\tilde{y}-\tilde{X}\beta)\sim WS(0, \tilde{\Sigma})$, an estimator $\hat{\gamma}(q\times 1)$ is the MMSLE based on $\tilde{y}$ of any linear vector function $\gamma=C\beta$ of β if and only if

(1) $\hat{\gamma}$ is a u-unbiased estimator of γ

and

(2) $\hat{\gamma}$ is a linear vector function

$$\hat{\gamma} = M\mathbf{g}$$

of the right-hand side $\mathbf{g}=(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{y})$ of the normal equations (2.7) where M is $q\times r$.

Proof: We show that $\hat{\gamma}$ is the unique MVLUE of γ. From Lemma 2.6 it follows that $\hat{\gamma}$ is the unique MMSLE of γ. To show the sufficiency of (1) and (2) let

$$\gamma^* = A\tilde{y}$$

be any other linear u-unbiased estimator of γ. Then

$$E(\gamma^*) = E(\gamma) = E(\hat{\gamma}),$$

from which it follows that

$$A\tilde{X}\mathbf{u} = C\mathbf{u} = M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X}\mathbf{u}. \tag{2.14}$$

Since (2.14) is an identity in $\mathbf{u}$ we have

$$(A\tilde{X} - M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X}) = 0, \tag{2.15}$$

$$(A\tilde{X} - C) = 0, \tag{2.16}$$

and

$$(M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X} - C) = 0. \tag{2.17}$$

Now

$$\begin{aligned}\operatorname{var}[\gamma^* - \gamma] &= \operatorname{var}[A\tilde{y} - C\beta]\\ &= \operatorname{var}[A(\tilde{X}\beta + \tilde{\varepsilon}) - C\beta]\\ &= \operatorname{var}[A\tilde{X}\mathbf{u} + A\tilde{X}u + A\tilde{\varepsilon} - C\mathbf{u} - Cu]\\ &= \operatorname{var}[(A\tilde{X} - C)u + A\tilde{\varepsilon}]\\ &= A\tilde{\Sigma}A'. \quad \text{(using (2.16))}\end{aligned}$$

Similarly

$$\begin{aligned}\operatorname{var}[\hat{\gamma} - \gamma] &= \operatorname{var}[M\tilde{X}'\tilde{\Sigma}^{-1}y - C\beta]\\ &= \operatorname{var}[M\tilde{X}'\tilde{\Sigma}^{-1}(\tilde{X}\beta + \tilde{\varepsilon}) - C\beta]\\ &= \operatorname{var}[M\tilde{X}'\tilde{\Sigma}^{-1}(\tilde{X}\mathbf{u} + \tilde{X}u + \tilde{\varepsilon}) - C\mathbf{u} - Cu]\\ &= \operatorname{var}[(M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X} - C)u + M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{\varepsilon}]\\ &= M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{\Sigma}\tilde{\Sigma}^{-1}\tilde{X}M' = M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X}M'.\end{aligned}$$

(using (2.17))

Write

$$\begin{aligned}&\operatorname{var}[\gamma^* - \gamma]\\ &= [A - M\tilde{X}'\tilde{\Sigma}^{-1} + M\tilde{X}'\tilde{\Sigma}^{-1}]\tilde{\Sigma}[A' - \tilde{\Sigma}^{-1}\tilde{X}M' + \tilde{\Sigma}^{-1}\tilde{X}M']\\ &= [A - M\tilde{X}'\tilde{\Sigma}^{-1}]\tilde{\Sigma}[A' - \tilde{\Sigma}^{-1}\tilde{X}M'] \qquad (2.18)\\ &\quad + 2[A - M\tilde{X}'\tilde{\Sigma}^{-1}]\tilde{X}M' + M\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X}M'\\ &= \text{non-negative quadratic form} + 0 + \operatorname{var}[\hat{\gamma} - \gamma]\end{aligned}$$

(using (2.15))

Since the diagonal elements of a non-negative quadratic form are non-negative,

$$\operatorname{var}[\gamma_i^* - \gamma_i] \geq \operatorname{var}[\hat{\gamma}_i - \gamma_i] \qquad i = 1, \cdots, q.$$

Thus $\hat{\gamma}$ is a MVLUE of γ. Uniqueness follows from (2.18), for if $\operatorname{var}[\gamma^*-\gamma]=\operatorname{var}[\hat{\gamma}-\gamma]$ it is necessary that $A=M\tilde{X}'\tilde{\Sigma}^{-1}$ and hence that $\gamma^*=\hat{\gamma}$. This establishes the sufficiency of conditions (1) and (2). Their necessity follows from Lemma 2.6 for (1) and a similar proof by contradiction for (2).

2.8 Corollary: The MMSLE of Random β. Given that

$$(\tilde{y} - \tilde{X}\beta) \sim WS(0, \tilde{\Sigma})$$

then

$$b = (\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}\tilde{X}'\tilde{\Sigma}^{-1}\tilde{y}$$

is the MMSLE of β based on $\tilde{y}$.

Proof: Note that (1) $E(b-\beta)=\mathbf{0}$ from Lemma 2.4, and (2) $b=M\mathbf{g}$ where $\mathbf{g}=\tilde{X}'\tilde{\Sigma}^{-1}\tilde{y}$, the right-hand side of the normal equations (2.7), and $M=(\tilde{X}'\tilde{\Sigma}^{-1}\tilde{X})^{-1}$. The desired conclusion follows from the Extended Gauss-Markov Theorem 2.7.

In Section 3 we shall review briefly a typical Kalman-type dynamic model discussed in the introduction and present the optimal estimators. In Section 4 we shall show how these optimal estimators may be derived from the random-β regression theory given above.

3. LINEAR DYNAMIC RECURSIVE ESTIMATION

3.1 The Model

A typical form of a linear dynamic model [5, 6] for a p-variate time series $y_1, \cdots, y_n$ is given by the recursive equations (using regression-like notation) as:

$$\beta_t = T_t\beta_{t-1} + u_t \tag{3.1}$$

$$y_t = X_t\beta_t + \varepsilon_t \tag{3.2}$$

where

$$\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} \sim WS\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q_t & 0 \\ 0 & R_t \end{bmatrix}\right). \tag{3.3}$$

for $t=1, \cdots, n$. The starting equation for β_t may be written

$$\beta_1 = \mu_1 + u_1 \tag{3.4}$$

where μ_1 is assumed known. All the error vectors are uncorrelated between times. Equation (3.1) expresses the unknown state $\beta_t(r\times 1)$ as a known linear transformation of the previous state β_{t-1} plus a vector u_t of random errors. As mentioned before, it is the linear dynamic state transition equation (3.1) in the parameters β_t which is new to conventional regression theory and gives the model so much of its versatility.

3.2 The Recursive Estimation Equation

In the engineering literature it is shown [5, 6] that the MMSLE b_t of β_t based on all the data $\mu_1, y_1, \cdots, y_t$ up through time t is given by the recursive *updating* equations:

$$b_t = b_{t|t-1} + S_{t|t-1}X_t' D_t^{-1}(y_t - X_t b_{t|t-1}) \tag{3.5a}$$

and

$$S_t = S_{t|t-1} - S_{t|t-1}X_t' D_t^{-1} X_t S_{t|t-1} \tag{3.5b}$$

where

$b_{1|0}=\mu_1$; $S_{1|0}=Q_1$;

$b_{t|t-1}=T_t b_{t-1}$; $S_{t|t-1}=T_t S_{t-1} T_t' + Q_t$; $\quad t=2, \cdots, n$

$D_t=R_t+X_t S_{t|t-1} X_t' \quad t=1, \cdots, n.$

It is also shown that

$$(b_t - \beta_t) \sim WS(0, S_t) \tag{3.6}$$

and that

$$E(b_t - \beta_t)\tilde{y}_t' = 0 \tag{3.7}$$

where

$$\tilde{y}_t = (\mu_1', y_1', \cdots, y_t')'.$$

Kalman [6] derives these results from the point of view of wide-sense conditional distributions and expectations as described, e.g., in [3]. In this approach b_t is obtained as an orthogonal projection of β_t on a linear manifold spanned by the elements of $\tilde{y}_t$. As mentioned previously, this is the same as fitting a regression of β_t on the elements of $\tilde{y}_t$ using *expected* first and second moments in the normal equations involved. Result (3.7) represents the special form of orthogonality so obtained, and the fact that $E(b_t-\beta_t)^2$ is minimized follows at the same time.

Another method of deriving similar stronger results is by the Bayesian approach under the Gaussian assumptions given in Section 1. The estimator b_t in (3.5a) is readily shown to be the posterior mean for β_t given $\tilde{y}_t$ and is thus the Bayes estimator under squared error loss or, in other words, the MMSE for β_t. For a complete discussion see [8].

The updating equations (3.5) are equivalent to the more intuitive recursive weighting equations [cf. (2.8) and (2.9)]:

$$b_t = S_t[S_{t|t-1}^{-1} b_{t|t-1} + X_t' R_t^{-1} y_t] \tag{3.8a}$$

and

$$S_t = [S_{t|t-1}^{-1} + X_t' R_t^{-1} X_t]^{-1}. \tag{3.8b}$$

Usually equations (3.5) are more convenient than the weighting equations (3.8) for computing, especially in the common cases $p<r$. The equivalence of equations (3.5) and (3.8) follows from the following two well-known lemmas.

Lemma 3.3: If $S=[M^{-1}+X'R^{-1}X]^{-1}$

then

$$S = M - MX'D^{-1}XM$$

where M is $r\times r$, R is $p\times p$, X is $p\times r$, and $D=R+XMX'$.

Proof: The result follows from showing that

$$\begin{aligned}
&[M^{-1} + X'R^{-1}X][M - MX'D^{-1}XM] \\
&= I - X'D^{-1}XM + X'R^{-1}XM - X'R^{-1}XMX'D^{-1}XM \\
&= I + [-X'D^{-1} + X'R^{-1} - X'R^{-1}XMX'D^{-1}]XM \\
&= I + X'R^{-1}[-RD^{-1} + I - XMX'D^{-1}]XM \\
&= I + X'R^{-1}[I - DD^{-1}]XM = I.
\end{aligned}$$

Lemma 3.4: If $b_1=S[M^{-1}b_0+X'R^{-1}y]$

then

$$b_1 = b_0 + MX'D^{-1}[y - \hat{y}]$$

where

M is $r \times r$, b_0 is $r \times 1$, X is $p \times r$, R is $p \times p$, y is $p \times 1$, $D=R+XMX'$, $\hat{y}=Xb_0$, and $S=[M^{-1}+X'R^{-1}X]^{-1}$.

Proof: Using the previous lemma we see that

$$\begin{aligned}
b_1 &= [M - MX'D^{-1}XM][M^{-1}b_0 + X'R^{-1}y] \\
&= b_0 + MX'[R^{-1}y - D^{-1}Xb_0 - D^{-1}XMX'R^{-1}y] \\
&= b_0 + MX'D^{-1}[(D - XMX')R^{-1}y - Xb_0] \\
&= b_0 + MX'D^{-1}[y - \hat{y}].
\end{aligned}$$

4. REGRESSION APPROACH DERIVATION

Using the random-β regression theory developed in Section 2, we derive the distribution of $b_t-\beta_t$ with b_t expressed in the intuitive weighting form (3.8a) and show directly that this b_t is the MMSLE of β_t based on all the data $\tilde{y}_t=[\mu_1', y_1', \cdots, y_t']'$ up through time t. Since equation (3.8a) is equivalent to equation (3.5a) we have an indirect derivation of the Kalman updating optimal estimator and its distribution.

Lemma 4.1: The distribution of $b_t-\beta_t$. Given the model described in equations (3.1) through (3.4), the recursive weighting equations (3.8a) and (3.8b) give b_t and S_t

such that

$$(b_t - \beta_t) \sim WS(0, S_t), \quad t = 1, \cdots, n. \tag{4.1}$$

Proof: To prove this inductively, suppose the result (4.1) is true for time $t-1$. Then from

$$b_{t|t-1} - \beta_t = T_t b_{t-1} - T_t \beta_{t-1} - u_t = T_t(b_{t-1} - \beta_{t-1}) - u_t$$

we have

$$(b_{t|t-1} - \beta_t) \sim WS(0, S_{t|t-1}). \tag{4.2}$$

Now letting

$$y_t^\dagger = X_t^\dagger \beta_t + \varepsilon_t^\dagger$$

represent

$$\begin{bmatrix} b_{t|t-1} \\ y_t \end{bmatrix} = \begin{bmatrix} I \\ X_t \end{bmatrix} \beta_t + \begin{bmatrix} b_{t|t-1} - \beta_t \\ \varepsilon_t \end{bmatrix} \tag{4.3}$$

we have, from (4.2)

$$(y_t^\dagger - X_t^\dagger \beta_t) \sim WS\left(0, \begin{bmatrix} S_{t|t-1} & 0 \\ 0 & R_t \end{bmatrix}\right). \tag{4.4}$$

Using (4.4) and the distribution Lemma 2.4(1) it is seen that the recursive weighting equations (3.8a) and (3.8b) give b_t and S_t such that $(b_t - \beta_t) \sim WS(0, S_t)$. This completes the induction step.

From the initial equations $b_{1|0} = \mu_1$ and $S_{1|0} = Q_1$, the premise (4.2) is satisfied at time $t=1$. Hence the Lemma is true at $t=1$ and therefore by induction at the successive times $t = 2, 3, \cdots, n$.

From model (4.4) above, the Extended Gauss-Markov Theorem 2.7 tells us that b_t is the MMSLE of β_t based on $b_{t|t-1}$ and y_t. It remains to be shown that b_t is MMSLE of β_t based on all the data $\tilde{y}_t$ up through time t. We show this in the following theorem.

Theorem 4.2: The Optimality of b_t. Given the recursive model described in equations (3.1) through (3.4), the recursive estimating equations (3.8a) and (3.8b) give the MMSLE b_t of β_t based on all the data $\mu_1, y_1, \cdots, y_t$ up through time t.

Proof: Write the recursive dynamic model (3.1) through (3.4) in the regression form

$$y_t^* = X_t^* \beta_t^* + \varepsilon_t^* \tag{4.5}$$

to represent

$$\begin{bmatrix} \mu_1 \\ y_1 \\ 0 \\ y_2 \\ \vdots \\ 0 \\ y_t \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \cdots 0 \\ X_1 & 0 & 0 \cdots 0 \\ -T_2 & I & 0 \cdots 0 \\ 0 & X_2 & 0 \cdots 0 \\ \vdots & & \\ 0 \cdots 0 & -T_t & I \\ 0 \cdots 0 & 0 & X_t \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_{t-1} \\ \beta_t \end{bmatrix} + \begin{bmatrix} -u_1 \\ \varepsilon_1 \\ -u_2 \\ \varepsilon_2 \\ \vdots \\ -u_t \\ \varepsilon_t \end{bmatrix}$$

where the third row partition, for example, comes from rewriting

$$\beta_2 = T_2 \beta_1 + u_2$$

as

$$0 = \beta_2 - T_2 \beta_1 - u_2.$$

Let

$$R_t^* \equiv \text{var}(\varepsilon_t^*).$$

Note that the right hand side of the normal equations for (4.5) is

$$g_t^* = X_t^{*\prime} R_t^{*-1} y_t^*$$

which reduces to

$$g_t^* = \begin{bmatrix} X_{t-1}^{*\prime} R_{t-1}^{-1*} y_{t-1}^* \\ X_t' R_t^{-1} y_t \end{bmatrix},$$

and that b_t is a linear function of g_t^*.

Since b_t is a linear function of the right-hand side of the normal equations for (4.5) and b_t is u-unbiased from Lemma 4.1, we conclude from the Extended Gauss-Markov Theorem that b_t is MMSLE of β_t based on all the data y_t^* and therefore $\tilde{y}_t$. This then establishes the optimality of b_t from the point of view of regressing y_t^* on X_t^*.

We have shown that b_t is also the MMSLE of β_t based on $b_{t|t-1}$ and y_t. In addition to the statistic g_t^*, the statistics $b_{t|t-1}$ and y_t may well be called wide-sense sufficient statistics for estimating any linear function $\gamma = C\beta$ of β just as $X'\Sigma^{-1}y$ is the complete and sufficient statistic for the fixed-β model $y = X\beta + \varepsilon$ where $\varepsilon \sim N(0, \Sigma)$.

We show the final property (3.7) in the following lemma.

Lemma 4.3: Given the conditions of Theorem 4.2,

$$E(b_t - \beta_t)\tilde{y}_t' = 0. \tag{4.6}$$

Proof: Using Lemma 2.4(3) in model (4.5)

$$E(b_t - \beta_t) y_t^{*\prime} = 0$$

and (4.6) follows from this.

5. CONCLUDING REMARKS

The conventional regressionist will have found our approach to the Kalman recursive results in Section 4 more natural since we derived them from the regression point of view. One of the keys to our approach is the fact that the recursive dynamic model (3.1) to (3.4) may be written in the form $y_t^* = X_t^* \beta_t^* + \varepsilon_t^*$, (4.5). We plan to show in future articles that this opens the way for the application of regression techniques to new problems arising in recursive estimation, e.g., the problem of estimating the R_t and Q_t variance matrices.

Before closing, it should be noted that Kalman's recursive estimation theory is considerably more general than the form discussed in this article. The more general problem is to find an optimal estimator b_t based on all the data $\tilde{y}_s$ up through time s: If $s<t$, $s=t$, or $s>t$, this is,

results can be obtained, e.g., that b_t is minimum mean square in the class of all (and not just linear) estimators.

For brevity we have dealt with only one single, but important, case of filtering under the wide-sense error assumptions. However, this illustrates a method for deriving similar results in the more general cases.

[*Received November 1971. Revised May 1972.*]

REFERENCES

[1] Bose, R.C., personal communication, 1956.

[2] Burnett, Thomas D. and Guthrie, Donald, "Estimation of *American Statistical Association*, 65 (December 1970), 1547–53.

Royal Statistical Society, Ser. B, 31 (No. 3, 1969), 446–54.

[5] Kalman, R.E., "New Methods in Wiener Filtering Theory," in J.L. Bodanoff and F. Kozin, eds., *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*, New York: John Wiley and Sons, Inc., 1963, 270–388.

[6] ———, "A New Approach to Linear Filtering and Prediction Problems," *Transactions ASME Journal of Basic Engineering*, 82 (March 1960), 35–45.

[7] Rao, C. Radhakrishna, *Linear Statistical Inference and its Applications*, New York: John Wiley and Sons, Inc., 1965, 192–3.

[8] Sage, Andrew P., *Optimum Systems Control*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968, 265–75.

Leondis, ed., *Advances in Control Systems*, 3, New York: Academic Press, 1966, 219–92.